ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

J. Math. Anal. Appl. 315 (2006) 225–243

Journal of MATHEMATICAL ANALYSIS AND APPLICATIONS

www.elsevier.com/locate/jmaa

Uniform recursive trees: Branching structure and simple random downward walk ☆

Chun Su *, Qunqiang Feng, Zhishui Hu

Department of Statistics and Finance, University of Science and Technology of China, Hefei 230026, China

Received 4 April 2004

Available online 25 May 2005

Submitted by J. Glaz

Abstract

As models for spread of epidemics, family trees, etc., various authors have used a random tree called the uniform recursive tree. Its branching structure and the length of simple random downward walk (SRDW) on it are investigated in this paper. On the uniform recursive tree of size n, we first give the distribution law of $\zeta_{n,m}$, the number of m-branches, whose asymptotic distribution is the Poisson distribution with parameter $\lambda = \frac{1}{m}$. We also give the joint distribution of the numbers of various branches and their covariance matrix. On L_n, the walk length of SRDW, we first give the exact expression of $\mathrm{P}(L_n = 2)$. Finally, the asymptotic behavior of L_n is given.

Keywords: Uniform recursive tree; Branch; SRDW; Walk length; Asymptotic behavior

1. Introduction

A tree is a simple connected graph without cycles [16]. The recursive tree of size n is a kind of random trees on n particles that attach to each other randomly. The process of

☆ This work was supported by the National Natural Science Foundation (Grant No. 10371117) and Special Foundation of USTC.

* Corresponding author.

E-mail addresses: suchun@ustc.edu.cn (C. Su), fengqq@mail.ustc.edu.cn (Q. Feng), huzs@mail.ustc.edu.cn (Z. Hu).

0022-247X/$ – see front matter
doi:10.1016/j.jmaa.2005.05.004

generating a recursive tree is as follows (see [2]): let the set of particles be $\{1, 2, \dots, n\}$, and $\{p_{k,i}, i = 1, 2, \dots, k\}$, $k = 1, 2, \dots, n-1$, be a sequence of probability mass functions, i.e.,

$$p_{k,i} \geqslant 0, \qquad \sum_{i=1}^{k} p_{k,i} = 1, \quad k = 1, 2, \dots, n-1.$$

At step 1, put all particles in a plane; at step 2, particle 2 attaches to particle 1; at step 3, particle 3 attaches to particle 1 with probability p_{21} or to particle 2 with probability p_{22}. In general, at step $k+1$, particle $k+1$ attaches one of the particles in the set $\{1, 2, \dots, k\}$ with the probabilities $p_{k,i}$, $i = 1, 2, \dots, k$, respectively. After n steps, the resulting tree with the root vertex 1 is called a recursive tree. If

$$p_{k,i} = \frac{1}{k}, \quad i = 1, 2, \dots, k, \ k = 1, 2, \dots, n-1,$$

i.e., at each step the new particle attaches to a uniformly selected particle from the previous ones, independently of previous attachments, then we call it a uniform recursive tree, denoted by $\mathcal{T}_n$. For any nature number $k \geqslant 2$, at the kth ($k \geqslant 2$) step we can make $k-1$ choices, so $(n-1)!$ different trees can be obtained, and each tree occurs with the same probability $\frac{1}{(n-1)!}$.

With many applications, recursive trees have been proposed as models for the spread of epidemics [13], the family trees of preserved copies of ancient or medieval texts [14], and pyramid schemes [5], etc. Here we give an example of the model for the spread of epidemics:

Example 1.1. Suppose there exists n persons infected a specific infectious disease (e.g., SARS) in turn in some area, and only one of them is the original case. The second case must be infected by the original one. Unknowing the law of infection, we suppose that the third case was infected by one of the previous two with the probability $1/2$, respectively. In general, we suppose the kth case was infected by one of the previous $k-1$ cases with respective probabilities $\frac{1}{k-1}$, $k = 2, 3, \dots, n$. Let vertex k represent the kth case, and vertex i attaches to vertex j ($1 \leqslant i < j \leqslant n$) if and only if the jth case was infected by the ith case. Then we obtain a uniform recursive tree. By this taken, such a study of study uniform recursive trees can make the law of infection clear to a certain extent.

In $\mathcal{T}_n$, D_j denotes the set of vertices of the jth generation. A subtree with the root in D_1 is called a branch, which is also a uniform recursive tree [12]. Obviously, the number of branches is $|D_1|$, denoted by η_n. If the size of a branch is m ($1 \leqslant m \leqslant n-1$), we call it an m-branch, and let $\zeta_{n,m}$ denote the number of the m-branches. In particular, if $m = 1$, the only vertex in the branch is called a child-leaf of the root 1. It is easy to see that $\eta_n = \sum_{i=1}^{n-1} \zeta_{n,i}$. Furthermore, if vertex $k \in D_j$, we say that the depth of vertex k is j, and let ξ_k denote the depth of vertex k.

Many authors have studied the depth of vertices. For example, Szymański [15] has given the distribution of ξ_n, the depth of vertex n; Devroye [3] has proved the central limit theorem of ξ_n; Mahmoud [9,10] has done some further study on the limiting behavior of

ξ_n and $\sum_{k=1}^{n} \xi_k$; Meir and Moon [12] have given the distribution of the number of vertices in each generation.

It is easy to see that the branching structure is one of the important properties of the uniform recursive trees, but as far as we know that no one has considered it. For each nature number n and $1 \leqslant m \leqslant n-1$, we shall show that

$$\mathrm{P}(\zeta_{n,m} = k) = \frac{1}{m^k k!} \sum_{i=0}^{[\frac{n-1}{m}]-k} \frac{(-1)^i}{i!m^i}, \quad k = 0, 1, \ldots, \left[\frac{n-1}{m}\right],$$

where $[t]$ is the biggest integer not more than t. Then for each fixed m,

$$\mathrm{P}(\zeta_{n,m} = k) \sim \mathrm{Poisson}\left(\frac{1}{m}\right),$$

as $n \to \infty$, where Poisson(λ) denotes the Poisson distribution with the parameter λ.

As well as the structure of the branches in uniform recursive trees, we have studied a random walk on them. Let G be a finite or infinite undirected graph with numerical vertex labels and a designated initial vertex s. By a local search on G, we mean the following. Place a particle on the initial vertex s. Examine the neighbors of s in turn until a vertex s' with a higher label is discovered. (If none exists, the process terminates.) Then move the particle to s' and continue. An evolutionary walk on a tree is local search beginning at the root [1]. The evolutionary walk is a stochastic process, which has been used to model local search in combinatorial optimization and molecular evolution. It was first correctly analyzed by Macken, Hagan and Perelson. Most of their results can be founded in [6–8].

Simple random downward walk (SRDW) on uniform recursive trees defined as follows is a special case of the evolutionary walks: beginning at the root 1, the particle moves to one of the children with the uniform probability, until the process terminates on some leaf. Throughout this paper we assume that $n \geqslant 2$. The walk length of SRDW, denoted by L_n, is defined to be the number of times the particle is moved, including its first placement. Then L_n is a random variable which takes values on $\{2, 3, \ldots, n\}$. Using the method of generating function, Meir and Moon [11] have shown that for each nature number $k \geqslant 2$,

$$\mathrm{P}(L_n = k) \sim \frac{(\log\log n)^{k-2}}{(k-2)!\log n}, \tag{1.1}$$

as $n \to \infty$. In this paper, based on the branching structure and using the method of probability, we have given the exact expression for the case $k = 2$ and an alternative proof of the relation (1.1), which has shown the inner relations between the random walk and the branching structure. For each n, as we shall show that

$$\mathrm{P}(L_n = 2) = \mathrm{E}\frac{\zeta_{n,1}}{\eta_n}$$

and for $3 \leqslant k \leqslant n-1$,

$$\mathrm{P}(L_n = k) = \sum_{m=k-1}^{n-1} \mathrm{P}(L_m = k-1)\mathrm{E}\frac{\zeta_{n,m}}{\eta_n}.$$

In the model for the spread of epidemics, $(L_n = k)$ represents the random event that k persons were infected in turn and none else was infected by the last case.

The results in this paper are organized as follows. In Section 2, we give the distribution, the asymptotic distribution and joint distribution of the branches' number in $\mathcal{T}_n$. Based on them, we come to study $\mathrm{P}(L_n = k)$. In Section 3, we give the exact expression of $\mathrm{P}(L_n = 2)$ and point out that it is strictly decreasing in n. In Section 4, using the method of probability, we newly proved the expression (1.1).

2. Branches of uniform recursive trees

To study L_n, we consider the properties of the branches first.

2.1. Distribution and asymptotic distribution of the branches' number

Obviously, for any $m \in \{1, \dots, n-1\}$, we have

$$\mathrm{P}(\zeta_{n,m} \geqslant 0) = 1, \qquad \mathrm{P}\left(\zeta_{n,m} > \left[\frac{n-1}{m}\right]\right) = 0. \tag{2.1}$$

Now we prove the following proposition.

Proposition 2.1. *In uniform recursive trees of size n, the distribution law of $\zeta_{n,m}$, the number of the m-branches, is the following*:

$$\mathrm{P}(\zeta_{n,m} = k) = \frac{1}{m^k k!} \sum_{i=0}^{[\frac{n-1}{m}]-k} \frac{(-1)^i}{i! m^i}, \quad k = 0, 1, \dots, \left[\frac{n-1}{m}\right]. \tag{2.2}$$

Specially, if $m > \frac{n-1}{2}$, $\zeta_{n,m}$ is a Bernoulli random variable, i.e.,

$$\mathrm{P}(\zeta_{n,m} = 1) = 1 - \mathrm{P}(\zeta_{n,m} = 0) = \frac{1}{m}.$$

Proof. From the set $\{2, 3, \dots, n\}$, i subsets of size m are chosen to make i m-branches (each may have $(m-1)!$ forms), and the rest of $n - mi - 1$ vertices attach arbitrarily by the rule above. Therefore, the number of the ways of generating a recursive tree is

$$\frac{\binom{n-1}{m}\binom{n-m-1}{m}\cdots\binom{n-m(i-1)-1}{m}((m-1)!)^i(n-mi-1)!}{i!} = \frac{(n-1)!}{m^i i!},$$
$$1 \leqslant i \leqslant \left[\frac{n-1}{m}\right]. \tag{2.3}$$

On the other hand, by (2.1),

$$\sum_{j=0}^{[\frac{n-1}{m}]} \mathrm{P}(\zeta_{n,m} = j) = 1.$$

Set $i = 1$ in (2.3), then

$$\frac{\binom{n-1}{m}(m-1)!(n-m-1)!}{(n-1)!} = \frac{1}{m}.$$

In view of the fact that each recursive tree which has j m-branches exactly is counted j times and $|\mathcal{T}_n| = (n-1)!$, the left of the equation above is $\sum_{j=1}^{[\frac{n-1}{m}]} \binom{j}{1} \mathrm{P}(\zeta_{n,m} = j)$. Thus,

$$\sum_{j=1}^{[\frac{n-1}{m}]} \binom{j}{1} \mathrm{P}(\zeta_{n,m} = j) = \frac{1}{m}. \tag{2.4}$$

Similarly,

$$\begin{aligned}
&\sum_{j=2}^{[\frac{n-1}{m}]} \binom{j}{2} \mathrm{P}(\zeta_{n,m} = j) = \frac{1}{2m^2}, \\
&\quad \ldots \\
&\sum_{j=i}^{[\frac{n-1}{m}]} \binom{j}{i} \mathrm{P}(\zeta_{n,m} = j) = \frac{1}{m^i i!}, \\
&\quad \ldots \\
&\mathrm{P}\left(\zeta_{n,m} = \left[\frac{n-1}{m}\right] - 1\right) + \left(\left[\frac{n-1}{m}\right]\right) \mathrm{P}\left(\zeta_{n,m} = \left[\frac{n-1}{m}\right]\right) \\
&\quad = \frac{1}{m^{[\frac{n-1}{m}]-1}\left([\frac{n-1}{m}] - 1\right)!}, \\
&\mathrm{P}\left(\zeta_{n,m} = \left[\frac{n-1}{m}\right]\right) = \frac{1}{m^{[\frac{n-1}{m}]}([\frac{n-1}{m}])!}.
\end{aligned} \tag{2.5}$$

Consider the $[\frac{n-1}{m}] + 1$ formulae above from the bottom up, then it is easy to yield (2.2).

If $m > \frac{n-1}{2}$, $\zeta_{n,m}$ only can take values of 0 or 1. Since $[\frac{n-1}{2}] = 1$,

$$\mathrm{P}(\zeta_{n,m} = 1) = 1 - \mathrm{P}(\zeta_{n,m} = 0) = \frac{1}{m}$$

follows by (2.2). $\square$

By (2.4) and (2.5), we can obtain the expectation and variance of $\zeta_{n,m}$.

Corollary 2.1.

(1) *For any* $n \geqslant 2$,

$$\mathrm{E}(\zeta_{n,m}) = \frac{1}{m}, \quad m = 1, \ldots, n-1.$$

(2) *For any* $n \geqslant 3$,

$$\mathrm{Var}(\zeta_{n,m}) = \begin{cases} \frac{1}{m}, & 1 \leqslant m \leqslant \frac{n-1}{2}; \\ \frac{m-1}{m^2}, & \frac{n-1}{2} < m \leqslant n-1. \end{cases} \tag{2.6}$$

Proof. It follows from (2.4) that $\mathrm{E}(\zeta_{n,m}) = \frac{1}{m}$. And by (2.5), if $1 \leqslant m \leqslant \frac{n-1}{2}$,

$$\mathrm{E}\big(\zeta_{n,m}^2\big) - \mathrm{E}(\zeta_{n,m}) = \frac{1}{m^2},$$

thus,

$$\mathrm{Var}(\zeta_{n,m}) = \mathrm{E}\big(\zeta_{n,m}^2\big) - \big(\mathrm{E}(\zeta_{n,m})\big)^2 = \frac{1}{m};$$

if $\frac{n-1}{2} < m \leqslant n-1$, the result is obvious. □

Let $\mathcal{N}$ be the set of nature numbers. From Proposition 2.1, it is easy to see that

Proposition 2.2. *For any $m \in \mathcal{N}$, the asymptotic distribution of $\zeta_{n,m}$ is the Poisson distribution with parameter $\lambda = \frac{1}{m}$, as $n \to \infty$.*

2.2. Joint distribution and numerical characteristics of the branches' numbers

Next we give the joint distribution of random vector $(\zeta_{n,1}, \zeta_{n,2}, \ldots, \zeta_{n,n-1})$.

Proposition 2.3. *In $\mathcal{T}_n$, the joint distribution of the numbers of various branches*

$$(\zeta_{n,1}, \zeta_{n,2}, \ldots, \zeta_{n,n-1})$$

is the following:

$$\mathrm{P}(\zeta_{n,1} = x_1, \zeta_{n,2} = x_2, \ldots, \zeta_{n,n-1} = x_{n-1}) = \prod_{m=1}^{n-1} \frac{1}{m^{x_m} x_m!}, \tag{2.7}$$

where $\{x_1, \ldots, x_{n-1}\}$ is any sequence of nonnegative integers satisfying the condition

$$\sum_{i=1}^{n-1} i x_i = n-1.$$

Proof. It suffices to compute the number of the elementary events (each one corresponds to a recursive tree) in the event $\{\zeta_{n,1} = x_1, \ldots, \zeta_{n,n-1} = x_{n-1}\}$. Consider the groups of $n-1$ vertices (vertices of a group belong to the same branch). The number of the ways of grouping is

$$\frac{(n-1)!}{(1!)^{x_1}(2!)^{x_2}\cdots((n-1)!)^{x_{n-1}}} \cdot \frac{1}{x_1!x_2!\cdots x_{n-1}!}.$$

And m-branch has $(m-1)!$ different forms, so the number of the elementary events in $\{\zeta_{n,1} = x_1, \ldots, \zeta_{n,n-1} = x_{n-1}\}$ is

$$\frac{(n-1)!}{(1!)^{x_1}(2!)^{x_2}\cdots((n-1)!)^{x_{n-1}}} \cdot \frac{(0!)^{x_1}(1!)^{x_2}\cdots[(n-2)!]^{x_{n-1}}}{x_1!x_2!\cdots x_{n-1}!}$$
$$= \frac{(n-1)!}{1^{x_1}2^{x_2}\cdots(n-1)^{x_{n-1}}x_1!x_2!\cdots x_{n-1}!}.$$

Since the elementary events occur with the same probability $\frac{1}{(n-1)!}$,

$$\begin{aligned}
&\mathrm{P}(\zeta_{n,1}=x_1,\zeta_{n,2}=x_2,\ldots,\zeta_{n,n-1}=x_{n-1})\\
&\quad=\frac{1}{(n-1)!}\cdot\frac{(n-1)!}{1^{x_1}2^{x_2}\cdots(n-1)^{x_{n-1}}x_1!x_2!\cdots x_{n-1}!}=\prod_{m=1}^{n-1}\frac{1}{m^{x_m}x_m!}. \qquad \square
\end{aligned}$$

In the previous subsection, we have obtained the expectation of random vector $(\zeta_{n,1}, \zeta_{n,2},\ldots,\zeta_{n,n-1})$, i.e.,

$$\mathrm{E}(\zeta_{n,1},\zeta_{n,2},\ldots,\zeta_{n,n-1})=\left(1,\frac{1}{2},\ldots,\frac{1}{n-1}\right). \tag{2.8}$$

Now we give its covariance matrix.

Proposition 2.4. *For any* $1\leqslant k<l\leqslant n-1$, *if* $k+l\leqslant n-1$,

$$\mathrm{Cov}(\zeta_{n,k},\zeta_{n,l})=0; \tag{2.9}$$

and if $k+l>n-1$,

$$\mathrm{Cov}(\zeta_{n,k},\zeta_{n,l})=-\frac{1}{kl}. \tag{2.10}$$

Proof. If $1\leqslant k<l\leqslant n-1$ and $k+l>n-1$, it is obvious that

$$\mathrm{P}(\zeta_{n,k}=i,\zeta_{n,l}=j)=0,\quad i,j>0,$$

thus,

$$\mathrm{E}(\zeta_{n,k}\zeta_{n,l})=0,\qquad \mathrm{Cov}(\zeta_{n,k},\zeta_{n,l})=-\mathrm{E}(\zeta_{n,k})\mathrm{E}(\zeta_{n,l})=-\frac{1}{kl}.$$

For any $1\leqslant k<l\leqslant n-1$ and $k+l\leqslant n-1$, if $ij>0, ik+jl\leqslant n-1$, the number of the uniform recursive trees which exactly have i k-branches and j l-branches is $(n-1)!\cdot \mathrm{P}(\zeta_{n,k}=i,\zeta_{n,l}=j)$. Let A and B be two disjoint subsets of the set $\{2,3,\ldots,n\}$, whose sizes are k and l, respectively. Then the number of uniform recursive trees, which have a k-branch and a l-branch consisting of the vertices in A and B, is $(k-1)!(l-1)!(n-k-l-1)!$. Noting that A and B can be chosen arbitrarily, the number multiplied by $\binom{n-1}{k}\cdot\binom{n-k-1}{l}$ is

$$M:=\binom{n-1}{k}\binom{n-k-1}{l}(k-1)!(l-1)!(n-k-l-1)!=\frac{(n-1)!}{kl}.$$

It is easy to see that in M, each recursive tree which exactly has i k-branches and j l-branches is counted ij times. Then

$$\sum_{(i,j):\ ij>0,\ ik+jl\leqslant n-1} ij(n-1)!\mathrm{P}(\zeta_{n,k}=i,\zeta_{n,l}=j)=M=\frac{(n-1)!}{kl}.$$

That is

$$\sum_{(i,j):\ ij>0,\ ik+jl\leqslant n-1} ij\mathrm{P}(\zeta_{n,k}=i,\zeta_{n,l}=j)=\frac{1}{kl}.$$

Hence, by (2.8),

$$\mathrm{E}(\zeta_{n,k}\zeta_{n,l}) = \frac{1}{k\cdot l} = \mathrm{E}(\zeta_{n,k})\mathrm{E}(\zeta_{n,l}), \quad 1 \leqslant k < l \leqslant n-1,\ k+l \leqslant n-1,$$

and $\mathrm{Cov}(\zeta_{n,k}, \zeta_{n,l}) = 0$ follows. $\square$

From this proposition, the following corollary is obvious.

Corollary 2.2. *The covariance matrix of random vector* $(\zeta_{n,1}, \zeta_{n,2}, \ldots, \zeta_{n,n-1})$ *is* $B_n = (b_{ij})_{(n-1)\times(n-1)}$, *where*

$$b_{ii} = \begin{cases} \frac{1}{i}, & 1 \leqslant i \leqslant \frac{n-1}{2}, \\ \frac{i-1}{i^2}, & \frac{n-1}{2} < i \leqslant n-1, \end{cases} \qquad b_{ij} = \begin{cases} 0, & i \neq j, i+j \leqslant n-1, \\ -\frac{1}{ij}, & i \neq j, i+j > n-1. \end{cases}$$

3. The exact expression of $\mathrm{P}(L_n = 2)$

From this section on, we shall discuss the probability on $\{L_n = k\}$. The main purpose in this section is to give the exact expression of $\mathrm{P}(L_n = 2)$. In the model for the spread epidemics, as described in the introduction, it means the ratio of the sufferers who do not infect others, in the group infected by the first sufferer. First, we give a elementary expression of it.

Proposition 3.1.

$$\mathrm{P}(L_n = 2) = \mathrm{E}\frac{\zeta_{n,1}}{\eta_n}. \tag{3.1}$$

Proof. By the rule of SRDW, conditioning on $\{\zeta_{n,1} = j, \eta_n = k\}$, the event $\{L_n = 2\}$ occurs with the probability $\frac{j}{k}$. According to total probability formula,

$$\begin{aligned}\mathrm{P}(L_n = 2) &= \sum_{1\leqslant j\leqslant k\leqslant n-1} \mathrm{P}(L_n = 2 \mid \zeta_{n,1} = j, \eta_n = k)\mathrm{P}(\zeta_{n,1} = j, \eta_n = k) \\ &= \sum_{1\leqslant j\leqslant k\leqslant n-1} \frac{j}{k}\mathrm{P}(\zeta_{n,1} = j, \eta_n = k) = \mathrm{E}\frac{\zeta_{n,1}}{\eta_n}.\end{aligned}$$

Acting similarly as above, the following consequence holds.

Corollary 3.1. *For any* $3 \leqslant k \leqslant n$,

$$\mathrm{P}(L_n = k) = \sum_{m=k-1}^{n-1} \mathrm{P}(L_m = k-1)\mathrm{E}\frac{\zeta_{n,m}}{\eta_n}. \tag{3.2}$$

The expression (3.1) relates to the joint distribution of $\zeta_{n,1}$ and η_n, which is hard to calculate. To get a expression of $\mathrm{P}(L_n = 2)$ which just relates to η_n, we give a lemma in the following.

Lemma 3.1. *For* $0 \leqslant j \leqslant k \leqslant n-1$ *and* $1 \leqslant m \leqslant n-1$, *we have*

$$\frac{1}{m(j+1)}\mathrm{P}(\zeta_{n,m}=j, \eta_n=k)=\mathrm{P}(\zeta_{n+m,m}=j+1, \eta_{n+m}=k+1).$$

Proof. Conditioning on that the event $\{\zeta_{n,m}=j\}$ occurs, it is easy to see that the size of the biggest branch of $\mathcal{T}_n$ is at most $n-1-mj$ or m. Using the joint distribution of $(\zeta_{n,1}, \zeta_{n,2}, \ldots, \zeta_{n,n-1})$ (see (2.7)), we have

$$\mathrm{P}(\zeta_{n,m}=j, \eta_n=k)=\sum_{\star}\frac{1}{m^j j!}\prod_{1\leqslant i\leqslant n-mj-1,\ i\neq m}\frac{1}{m^{x_m}x_m!},$$

where $\sum_\star$ is taken over all integers $\{x_1, x_2, \ldots, x_{m-1}, x_{m+1}, \ldots, x_{n-mj-1}\}$ satisfying the conditions of $x_1+x_2+\cdots+x_{m-1}+j+x_{m+1}+\cdots+x_{n-1-j}=k$ and $x_1+2x_2+\cdots+(m-1)x_{m-1}+mj+(m+1)x_{m+1}+\cdots+(n-mj-1)x_{n-mj-1}=n-1$. Thus,

$$\begin{aligned}\frac{1}{m(j+1)}\mathrm{P}(\zeta_{n,m}=j, \eta_n=k)&=\sum_{\star}\frac{1}{m^{j+1}(j+1)!}\prod_{1\leqslant i\leqslant n-mj-1,\ i\neq m}\frac{1}{m^{x_m}x_m!}\\&=\mathrm{P}(\zeta_{n+m,m}=j+1, \eta_{n+m}=k+1). \qquad \square\end{aligned}$$

The following theorem is one of our main results.

Theorem 3.1. *For any* $1 \leqslant m \leqslant n-1$, *we have*

$$\mathrm{E}\frac{\zeta_{n,m}}{\eta_n}=\frac{1}{m}\mathrm{E}\frac{1}{\eta_{n-m}+1}. \tag{3.3}$$

In particular,

$$\mathrm{P}(L_n=2)=\mathrm{E}\frac{\zeta_{n,1}}{\eta_n}=\mathrm{E}\frac{1}{\eta_{n-1}+1}=\frac{1}{(n-2)!}\int_0^1\prod_{j=0}^{n-3}(x+j)\,dx. \tag{3.4}$$

Proof. For any $1 \leqslant m \leqslant n-1$,

$$\begin{aligned}\frac{1}{m}\mathrm{E}\frac{1}{\eta_{n-m}+1}&=\frac{1}{m}\sum_{k=1}^{n-m-1}\frac{1}{k+1}\mathrm{P}(\eta_{n-m}=k)\\&=\sum_{k=1}^{n-m-1}\frac{1}{m(k+1)}\sum_{j=0}^{k}\mathrm{P}(\zeta_{n-m,m}=j, \eta_{n-m}=k)\\&=\sum_{0\leqslant j\leqslant k,\ 1\leqslant k\leqslant n-m-1}\frac{j+1}{k+1}\frac{1}{m(j+1)}\mathrm{P}(\zeta_{n-m,m}=j, \eta_{n-m}=k)\\&=\sum_{0\leqslant j\leqslant k,\ 1\leqslant k\leqslant n-m-1}\frac{j+1}{k+1}\mathrm{P}(\zeta_{n,m}=j+1, \eta_n=k+1)\\&=\sum_{1\leqslant i\leqslant l\leqslant n-m}\frac{i}{l}\mathrm{P}(\zeta_{n,m}=i, \eta_n=l)=\mathrm{E}\frac{\zeta_{n,m}}{\eta_n},\end{aligned}$$

and (3.3) holds. By Proposition 3.1, set $m = 1$ in this equation, then

$$\mathrm{P}(L_n = 2) = \mathrm{E}\frac{\zeta_{n,1}}{\eta_n} = \mathrm{E}\frac{1}{\eta_{n-1}+1}.$$

Since (see [3,9] or [4])

$$\mathrm{P}(\eta_n = k) = \frac{\beta_{n,n-k-1}}{(n-1)!}, \quad k = 1, \dots, n-1, \tag{3.5}$$

where

$$\beta_{n,0} = 1, \qquad \beta_{n,k} = \sum_{1 \leqslant m_1 \leqslant \cdots \leqslant m_k \leqslant n-2} m_1 \cdots m_k, \quad k = 1, \dots, n-2,$$

we have

$$\mathrm{P}(L_n = 2) = \mathrm{E}\frac{1}{\eta_{n-1}+1} = \frac{1}{(n-2)!}\sum_{k=1}^{n-2}\frac{\beta_{n-1,n-k-2}}{k+1} = \frac{1}{(n-2)!}\int_0^1 \prod_{j=0}^{n-3}(x+j)\,dx. \tag{3.6}$$

In the above expression, the sum is not only hard to calculate if n is large, but also hard to see how the probability $\mathrm{P}(L_n = 2)$ varies as $n \to \infty$. In the next section, we will make a further discussion.

Corollary 3.2. *The probability* $\mathrm{P}(L_n = 2)$ *is strictly decreasing in* n, *i.e.*,

$$\mathrm{P}(L_n = 2) > \mathrm{P}(L_{n+1} = 2), \quad n \geqslant 2.$$

Proof. It is easy to see that

$$\begin{aligned}\mathrm{P}(L_{n+1} = 2) &= \frac{1}{(n-1)!}\int_0^1 (x+n-2)\prod_{j=0}^{n-3}(x+j)\,dx \\ &< \frac{n-1}{(n-1)!}\int_0^1 \prod_{j=0}^{n-3}(x+j)\,dx = \mathrm{P}(L_n = 2). \qquad \square\end{aligned}$$

Obviously, the event $(L_n = k)$ can only occur on $\mathcal{T}_n$ of size $n \geqslant k$. It is easy to get the following:

$$\mathrm{P}(L_k = k) = \frac{1}{(k-1)!}, \qquad \mathrm{P}(L_{k+1} = k) = \frac{\binom{k+1}{2} - 1}{2 \cdot k!}.$$

If $k \geqslant 4$, then $k^2 - 3k - 2 > 0$ and $\mathrm{P}(L_{k+1} = k) > \mathrm{P}(L_k = k)$. It is shown that the probability $\mathrm{P}(L_n = k)$ is not strictly decreasing for $k \geqslant 4$.

4. The length of SRDW

As described in the introduction, we shall study the asymptotic behavior of the probability $\mathrm{P}(L_n = k)$, $k \geqslant 2$.

Theorem 4.1. *For any positive integer $k \geqslant 2$, we have*

$$\lim_{n\to\infty} \frac{(k-2)!\log n}{(\log\log n)^{k-2}} \cdot \mathrm{P}(L_n = k) = 1. \tag{4.1}$$

The probability $\mathrm{P}(L_n = 2)$ is the inductive base of the result, so we discuss its asymptotic behavior first.

4.1. The case $k = 2$

Proposition 4.1.

$$\lim_{n\to\infty} \log n \cdot \mathrm{P}(L_n = 2) = 1. \tag{4.2}$$

Proof. Write

$$\mathrm{P}(L_n = 2) = \frac{1}{(n-2)!} \int_0^1 \prod_{j=0}^{n-3} (x+j)\,dx = \int_0^1 \prod_{j=1}^{n-2} \left(1 + \frac{x-1}{j}\right) dx.$$

It is easy to see that

$$1 + t \leqslant e^t, \quad t > -1;$$
$$1 + t \geqslant e^{t-t^2}, \quad -\frac{1}{2} < t < 0.$$

Then

$$\prod_{j=1}^{n-2} \left(1 + \frac{x-1}{j}\right) \leqslant \exp\left\{(x-1)\sum_{j=1}^{n-2} \frac{1}{j}\right\} = \exp\left\{-(1-x)\sum_{j=1}^{n-2} \frac{1}{j}\right\}$$

and

$$\begin{aligned} \log n \cdot \mathrm{P}(L_n = 2) &\leqslant \log n \cdot \int_0^1 \exp\left\{-(1-x)\sum_{j=1}^{n-2}\frac{1}{j}\right\} dx \\ &= \log n \cdot \int_0^1 \exp\left\{-x\sum_{j=1}^{n-2}\frac{1}{j}\right\} dx \\ &= \frac{\log n}{\sum_{j=1}^{n-2}\frac{1}{j}} \left(1 - \exp\left\{-\sum_{j=1}^{n-2}\frac{1}{j}\right\}\right) \to 1, \quad n \to \infty. \end{aligned} \tag{4.3}$$

On the other hand, if $\frac{1}{2} < x < 1$, then $1 - x > (1-x)^2$ and

$$\prod_{j=1}^{n-2}\left(1+\frac{x-1}{j}\right) \geqslant \exp\left\{-(1-x)\sum_{j=1}^{n-2}\frac{1}{j}-(1-x)^2\sum_{j=1}^{n-2}\frac{1}{j^2}\right\}$$
$$\geqslant \exp\left\{-(1-x)\sum_{j=1}^{n-2}\frac{j+1}{j^2}\right\}.$$

Therefore,

$$\log n\cdot \mathrm{P}(L_n=2) \geqslant \log n\cdot\int_{\frac{1}{2}}^{1}\exp\left\{-(1-x)\sum_{j=1}^{n-2}\frac{j+1}{j^2}\right\}dx$$
$$=\log n\cdot\int_{0}^{\frac{1}{2}}\exp\left\{-x\sum_{j=1}^{n-2}\frac{j+1}{j^2}\right\}dx$$
$$=\frac{\log n}{\sum_{j=1}^{n-2}\frac{j+1}{j^2}}\left(1-\exp\left\{-\frac{1}{2}\sum_{j=1}^{n-2}\frac{j+1}{j^2}\right\}\right)\to 1,\quad n\to\infty. \tag{4.4}$$

By (4.3) and (4.4), Proposition 4.1 holds. □

From Corollary 3.1, we can write out the exact expressions for general cases, which are too complex to utilize. Using the above method, it is hard to get the asymptotic distributions of $\mathrm{P}(L_n=k)$ for $k=3,4,\ldots,$ so we shall use another method. First we give a lemma:

Lemma 4.1. *For any* $0<\varepsilon<1$ *and slowly varying sequence* $l(n)$ *as* $n\to\infty$,

$$\lim_{n\to\infty} l(n)\mathrm{P}\left(\left|\frac{\eta_n-\log n}{\log n}\right|\geqslant\varepsilon\right)=0.$$

In particular,

$$\lim_{n\to\infty}\log n\cdot \mathrm{P}\big(\eta_n\leqslant(1-\varepsilon)\log n\big)=0; \tag{4.5}$$
$$\lim_{n\to\infty}\log n\cdot \mathrm{P}\big(\eta_n\geqslant(1+\varepsilon)\log n\big)=0. \tag{4.6}$$

Proof. Let $X_j=I(j+1\in D_1)$, then (see [4])

$$\eta_n=\sum_{j=1}^{n-1}X_j,\quad n\geqslant 2,$$

where $X_1,X_2,\ldots,X_{n-1}$ are mutually independent Bernoulli random variables satisfying

$$\mathrm{P}(X_j=1)=1-\mathrm{P}(X_j=0)=\frac{1}{j},\quad j\geqslant 1.$$

Therefore,

$$\mathrm{E}e^{tX_j} = \left(1-\frac{1}{j}\right)+\frac{1}{j}e^t = 1+\frac{1}{j}\left(e^t-1\right) \leqslant \exp\left\{\frac{1}{j}\left(e^t-1\right)\right\}, \quad \forall t\in\mathcal{R},\ j\geqslant 1. \tag{4.7}$$

In the following, C shall denote an absolute constant, possibly varying from place to place. Hence, for any $t>0$,

$$\begin{aligned} \mathrm{P}\big(\eta_n \geqslant (1+\varepsilon)\log n\big) &\leqslant \exp\left\{-t(1+\varepsilon)\log n + \sum_{j=1}^{n-1}\frac{1}{j}\left(e^t-1\right)\right\} \\ &\leqslant \exp\left\{C\left(e^t-1\right)\right\}\exp\left\{\log n\cdot\left(e^t-1-(1+\varepsilon)t\right)\right\}. \end{aligned}$$

Choose a $t_1\in(0,\log(1+\varepsilon))$, then $\alpha_1 := -(e^{t_1}-1-(1+\varepsilon)t_1) > 0$ and

$$\mathrm{P}\big(\eta_n \geqslant (1+\varepsilon)\log n\big) \leqslant \exp\left\{C\left(e^{t_1}-1\right)\right\}n^{-\alpha_1}. \tag{4.8}$$

Similarly, for any $t>0$,

$$\begin{aligned} \mathrm{P}\big(\eta_n \leqslant (1-\varepsilon)\log n\big) &= \mathrm{P}(\log n - \eta_n \geqslant \varepsilon\log n) \\ &\leqslant \exp\left\{t(1-\varepsilon)\log n + \sum_{j=1}^{n-1}\frac{1}{j}\left(e^{-t}-1\right)\right\} \\ &\leqslant \exp\left\{C\left(e^{-t}-1\right)\right\}\exp\left\{\log n\cdot\left(e^{-t}-1+(1-\varepsilon)t\right)\right\} \\ &\leqslant \exp\left\{\log n\cdot\left(e^{-t}-1+(1-\varepsilon)t\right)\right\}. \end{aligned}$$

Choose a $t_2>0$ such that $\alpha_2 := -(e^{-t_2}-1+(1-\varepsilon)t_2) > 0$, then

$$\mathrm{P}\big(\eta_n \leqslant (1-\varepsilon)\log n\big) \leqslant n^{-\alpha_2}. \tag{4.9}$$

By (4.8) and (4.9), for any $\varepsilon>0$ and slowly varying sequence $l(n)$ as $n\to\infty$,

$$\lim_{n\to\infty} l(n)\mathrm{P}\left(\left|\frac{\eta_n-\log n}{\log n}\right| \geqslant \varepsilon\right) = 0.$$

Specially, take $l(n)=\log n$, then (4.5) and (4.6) holds. □

Using Lemma 4.1, we can prove Proposition 4.1 in another way.

An alternative proof of Proposition 4.1. It is easy to see that

$$\zeta_{n,1} \leqslant \eta_n \leqslant n-1. \tag{4.10}$$

According to (3.1), to prove the proposition, it suffices to prove that

$$\lim_{n\to\infty} \mathrm{E}\frac{\log n\cdot\zeta_{n,1}}{\eta_n} = 1. \tag{4.11}$$

Note that (see [9] or [4])

$$\mathrm{E}(\zeta_{n,1}) = \mathrm{Var}(\zeta_{n,1}) = 1, \quad \forall n\geqslant 3; \qquad \frac{\eta_n}{\log n} \xrightarrow{p} 1, \quad n\to\infty.$$

For any $\varepsilon>0$,

$$
\begin{aligned}
1 &= \mathrm{E}(\zeta_{n,1}) = \mathrm{E}\big(\zeta_{n,1} I\big(\eta_n \geqslant (1+\varepsilon)\log n\big)\big) + \mathrm{E}\big(\zeta_{n,1} I\big(\eta_n < (1+\varepsilon)\log n\big)\big) \\
&:= I_{n,1} + I_{n,2}.
\end{aligned}
\tag{4.12}
$$

Then $\mathrm{E}(\zeta_{n,1}^2) = \mathrm{Var}(\zeta_{n,1}) + (\mathrm{E}(\zeta_{n,1}))^2 = 2$, for all $n \geqslant 3$, and

$$
I_{n,1} \leqslant \sqrt{\mathrm{E}\big(\zeta_{n,1}^2\big)\mathrm{P}\big(\eta_n \geqslant (1+\varepsilon)\log n\big)} \leqslant \sqrt{2 \cdot \mathrm{P}\left(\left|\frac{\eta_n}{\log n} - 1\right| \geqslant \varepsilon\right)} \to 0, \quad n \to \infty.
$$

Thus,

$$
\lim_{n\to\infty} I_{n,2} = \lim_{n\to\infty} \mathrm{E}\big(\zeta_{n,1} I\big(\eta_n < (1+\varepsilon)\log n\big)\big) = 1.
\tag{4.13}
$$

Similarly,

$$
\lim_{n\to\infty} \mathrm{E}\big(\zeta_{n,1} I\big(\eta_n \geqslant (1-\varepsilon)\log n\big)\big) = 1.
\tag{4.14}
$$

By (4.13),

$$
\begin{aligned}
\liminf_{n\to\infty} \mathrm{E}\frac{\log n \cdot \zeta_{n,1}}{\eta_n} &\geqslant \liminf_{n\to\infty} \mathrm{E}\left(\frac{\log n \cdot \zeta_{n,1}}{\eta_n} I\big(\eta_n < (1+\varepsilon)\log n\big)\right) \\
&\geqslant \frac{1}{1+\varepsilon} \lim_{n\to\infty} \mathrm{E}\big(\zeta_{n,1} I\big(\eta_n < (1+\varepsilon)\log n\big)\big) = \frac{1}{1+\varepsilon}.
\end{aligned}
$$

Therefore,

$$
\liminf_{n\to\infty} \mathrm{E}\frac{\log n \cdot \zeta_{n,1}}{\eta_n} \geqslant 1,
\tag{4.15}
$$

from the arbitrariness of $\varepsilon > 0$.

On the other hand, for any $\varepsilon > 0$, write

$$
\begin{aligned}
\mathrm{E}\frac{\log n \cdot \zeta_{n,1}}{\eta_n} &= \mathrm{E}\left(\frac{\log n \cdot \zeta_{n,1}}{\eta_n} I\big(\eta_n < (1-\varepsilon)\log n\big)\right) \\
&\quad + \mathrm{E}\left(\frac{\log n \cdot \zeta_{n,1}}{\eta_n} I\big(\eta_n \geqslant (1-\varepsilon)\log n\big)\right) \\
&:= J_{n,1} + J_{n,2}.
\end{aligned}
\tag{4.16}
$$

By (4.14),

$$
\begin{aligned}
\limsup_{n\to\infty} J_{n,2} &= \limsup_{n\to\infty} \mathrm{E}\left(\frac{\log n \cdot \zeta_{n,1}}{\eta_n} I\big(\eta_n \geqslant (1-\varepsilon)\log n\big)\right) \\
&\leqslant \frac{1}{1-\varepsilon} \lim_{n\to\infty} \mathrm{E}\big(\zeta_{n,1} I\big(\eta_n \geqslant (1-\varepsilon)\log n\big)\big) = \frac{1}{1-\varepsilon}.
\end{aligned}
\tag{4.17}
$$

Since $\zeta_{n,1} \leqslant \eta_n$,

$$
\lim_{n\to\infty} J_{n,1} \leqslant \lim_{n\to\infty} \log n \cdot \mathrm{P}\big(\eta_n < (1-\varepsilon)\log n\big) = 0
\tag{4.18}
$$

follows by Lemma 4.1. And by (4.16)–(4.18),

$$
\limsup_{n\to\infty} \mathrm{E}\frac{\log n \cdot \zeta_{n,1}}{\eta_n} \leqslant \frac{1}{1-\varepsilon}.
$$

Thus,

$$\limsup_{n\to\infty} \mathrm{E}\frac{\log n \cdot \zeta_{n,1}}{\eta_n} \leqslant 1, \tag{4.19}$$

also from the arbitrariness of $\varepsilon > 0$. Then (4.1) follows by (4.19) and (4.15). □

An interesting consequence of Proposition 4.1 is as follows.

Corollary 4.1.

$$\mathrm{E}\frac{1}{\eta_n} \sim \frac{1}{\log n}, \quad n \to \infty. \tag{4.20}$$

Proof. To prove (4.20), it suffices to show that

$$\lim_{n\to\infty} \log n \left| \mathrm{E}\frac{1}{\eta_n} - \mathrm{E}\frac{1}{\eta_n + 1} \right| = 0. \tag{4.21}$$

Since $\eta_n \geqslant 1$,

$$\begin{aligned}
&\log n \left| \mathrm{E}\frac{1}{\eta_n} - \mathrm{E}\frac{1}{\eta_n + 1} \right| \\
&\quad = \left| \mathrm{E}\frac{\log n}{\eta_n(\eta_n + 1)} \right| \\
&\quad \leqslant \left| \mathrm{E}\frac{\log n}{\eta_n(\eta_n + 1)} I\left(\eta_n \geqslant \frac{1}{2}\log n\right) \right| + \left| \mathrm{E}\frac{\log n}{\eta_n(\eta_n + 1)} I\left(\eta_n < \frac{1}{2}\log n\right) \right| \\
&\quad \leqslant 2\mathrm{E}\frac{1}{\eta_n + 1} + \log n \cdot \mathrm{P}\left(\eta_n < \frac{1}{2}\log n\right).
\end{aligned}$$

Thus, by (3.4), (4.5) and Proposition 4.1, (4.21) follows. □

4.2. *The general cases*

Now we give our proof of the expression (1.1).

Proof of Theorem 4.1. By Corollary 3.1, to prove the theorem, it suffices to prove that

$$\lim_{n\to\infty} \frac{(k-2)!\log n}{(\log\log n)^{k-2}} \cdot \sum_{m=k-1}^{n-1} \mathrm{P}(L_m = k-1)\mathrm{E}\frac{\zeta_{n,m}}{\eta_n} = 1. \tag{4.22}$$

The result for $k = 2$ has been proved. Suppose that the equation above holds for $k - 1$ ($k \geqslant 3$). We shall prove it still holds for k in the following.

First we show that

$$\limsup_{n\to\infty} \frac{(k-2)!\log n}{(\log\log n)^{k-2}} \cdot \sum_{m=k-1}^{n-1} \mathrm{P}(L_m = k-1)\mathrm{E}\frac{\zeta_{n,m}}{\eta_n} \leqslant 1. \tag{4.23}$$

In fact, for any $0 < \varepsilon < 1$, write

$$\begin{aligned}
&\frac{(k-2)!\log n}{(\log\log n)^{k-2}}\cdot\sum_{m=k-1}^{n-1}\mathrm{P}(L_m=k-1)\mathrm{E}\frac{\zeta_{n,m}}{\eta_n}\\
&\quad=\frac{(k-2)!\log n}{(\log\log n)^{k-2}}\cdot\sum_{m=k-1}^{n-1}\mathrm{P}(L_m=k-1)\mathrm{E}\biggl(\frac{\zeta_{n,m}}{\eta_n}I\bigl(\eta_n>(1-\varepsilon)\log n\bigr)\biggr)\\
&\qquad+\frac{(k-2)!\log n}{(\log\log n)^{k-2}}\cdot\sum_{m=k-1}^{n-1}\mathrm{P}(L_m=k-1)\mathrm{E}\biggl(\frac{\zeta_{n,m}}{\eta_n}I\bigl(\eta_n\leqslant(1-\varepsilon)\log n\bigr)\biggr)\\
&\quad:=I_1(n)+I_2(n).
\end{aligned}\tag{4.24}$$

In view of the fact that $\mathrm{P}(L_m=k-1)\leqslant 1$ and $\eta_n=\sum_{m=1}^{n-1}\zeta_{n,m}$,

$$\begin{aligned}
I_2(n)&\leqslant\frac{(k-2)!\log n}{(\log\log n)^{k-2}}\mathrm{E}\Biggl(\Biggl(\frac{1}{\eta_n}\sum_{m=k-1}^{n-1}\mathrm{P}(L_m=k-1)\zeta_{n,m}\Biggr)I\bigl(\eta_n\leqslant(1-\varepsilon)\log n\bigr)\Biggr)\\
&\leqslant(k-2)!\log n\cdot\mathrm{P}\bigl(\eta_n\leqslant(1-\varepsilon)\log n\bigr).
\end{aligned}$$

Hence,

$$\lim_{n\to\infty}I_2(n)=0$$

by Lemma 4.1. And by (2.8) and the inductive assumption, we have

$$\begin{aligned}
I_1(n)&<\frac{1}{1-\varepsilon}\frac{(k-2)!}{(\log\log n)^{k-2}}\mathrm{E}\Biggl(\sum_{m=k-1}^{n-1}\mathrm{P}(L_m=k-1)\zeta_{n,m}I\bigl(\eta_n>(1-\varepsilon)\log n\bigr)\Biggr)\\
&\leqslant\frac{1}{1-\varepsilon}\frac{(k-2)!}{(\log\log n)^{k-2}}\cdot\sum_{m=k-1}^{n-1}\mathrm{P}(L_m=k-1)\mathrm{E}\zeta_{n,m}.
\end{aligned}$$

Since

$$\lim_{m\to\infty}\frac{\mathrm{P}(L_m=k-1)(k-3)!\log m}{(\log\log m)^{k-3}}=1$$

and

$$\lim_{n\to\infty}\frac{1}{(\log\log n)^{k-2}}\sum_{m=k-1}^{n-1}\frac{(k-2)(\log\log m)^{k-3}}{m\log m}=1,$$

it is easy to know

$$\begin{aligned}
&\lim_{n\to\infty}\frac{k-2}{(\log\log n)^{k-2}}\cdot\sum_{m=k-1}^{n-1}\mathrm{P}(L_m=k-1)\mathrm{E}\zeta_{n,m}\\
&\quad=\lim_{n\to\infty}\frac{k-2}{(\log\log n)^{k-2}}\cdot\sum_{m=k-1}^{n-1}\frac{\mathrm{P}(L_m=k-1)(k-3)!\log m}{(\log\log m)^{k-3}}\cdot\frac{(\log\log m)^{k-3}}{m\log m}\\
&\quad=1.
\end{aligned}\tag{4.25}$$

Thus, from the above, for any $0 < \varepsilon < 1$,

$$\limsup_{n\to\infty} \frac{(k-2)!\log n}{(\log\log n)^{k-2}} \cdot \sum_{m=k-1}^{n-1} \mathrm{P}(L_m = k-1)\mathrm{E}\frac{\zeta_{n,m}}{\eta_n} \leqslant \frac{1}{1-\varepsilon},$$

which yields (4.23) from the arbitrariness of $\varepsilon > 0$.

To prove the theorem, we need only to prove that

$$\liminf_{n\to\infty} \frac{(k-2)!\log n}{(\log\log n)^{k-2}} \cdot \sum_{m=k-1}^{n-1} \mathrm{P}(L_m = k-1)\mathrm{E}\frac{\zeta_{n,m}}{\eta_n} \geqslant 1. \tag{4.26}$$

First, for any $\varepsilon > 0$, we have

$$\lim_{n\to\infty} \frac{1}{(\log\log n)^{k-2}} \cdot \sum_{m=k-1}^{n-1} \mathrm{P}(L_m = k-1)\mathrm{E}\big(\zeta_{n,m} I\big(\eta_n > (1+\varepsilon)\log n\big)\big) = 0. \tag{4.27}$$

In fact, by (2.8) and (4.25),

$$\lim_{n\to\infty} \frac{1}{(\log\log n)^{k-2}} \cdot \sum_{m=k-1}^{n-1} \mathrm{P}(L_m = k-1)\mathrm{E}\zeta_{n,m}\mathrm{P}\big(\eta_n > (1+\varepsilon)\log n\big) = 0.$$

Hence, to prove (4.27), it suffices to show that

$$\begin{aligned}&\lim_{n\to\infty} \frac{1}{(\log\log n)^{k-2}} \\ &\quad\times \sum_{m=k-1}^{n-1} \mathrm{P}(L_m = k-1)\mathrm{E}\big((\zeta_{n,m} - \mathrm{E}\zeta_{n,m}) I\big(\eta_n > (1+\varepsilon)\log n\big)\big) = 0.\end{aligned} \tag{4.28}$$

And by Lemma 2.2,

$$\mathrm{Cov}(\zeta_{n,i}, \zeta_{n,j}) \leqslant 0, \quad \forall i \neq j, \qquad \mathrm{Var}\,\zeta_{n,m} \leqslant \frac{1}{m},$$

thus,

$$\begin{aligned}&\frac{1}{(\log\log n)^{k-2}} \cdot \sum_{m=k-1}^{n-1} \mathrm{P}(L_m = k-1)\mathrm{E}\big(\big(\zeta_{n,m} - \mathrm{E}(\zeta_{n,m})\big) I\big(\eta_n > (1+\varepsilon)\log n\big)\big) \\ &\quad= \frac{1}{(\log\log n)^{k-2}} \\ &\qquad\times \mathrm{E}\left(\left(\sum_{m=k-1}^{n-1} \mathrm{P}(L_m = k-1)\big(\zeta_{n,m} - \mathrm{E}(\zeta_{n,m})\big)\right) I\big(\eta_n > (1+\varepsilon)\log n\big)\right) \\ &\quad\leqslant \frac{1}{(\log\log n)^{k-2}} \\ &\qquad\times \left(\mathrm{E}\left(\sum_{m=k-1}^{n-1} \mathrm{P}(L_m = k-1)\big(\zeta_{n,m} - \mathrm{E}(\zeta_{n,m})\big)\right)^2 \mathrm{P}\big(\big(\eta_n > (1+\varepsilon)\log n\big)\big)\right)^{1/2}\end{aligned}$$

$$\leqslant \frac{1}{(\log\log n)^{k-2}}\left(\left(\sum_{m=k-1}^{n-1} \mathrm{P}^2(L_m = k-1)\operatorname{Var}(\zeta_{n,m})\right)\mathrm{P}\big(\big(\eta_n > (1+\varepsilon)\log n\big)\big)\right)^{1/2}$$

$$\leqslant \frac{1}{(\log\log n)^{k-2}}\left(\left(\sum_{m=k-1}^{n-1} \frac{\mathrm{P}^2(L_m = k-1)}{m}\right)\mathrm{P}\big(\big(\eta_n > (1+\varepsilon)\log n\big)\big)\right)^{1/2}.$$

By the inductive assumption,

$$\sum_{m=k-1}^{\infty} \frac{\mathrm{P}^2(L_m = k-1)}{m} = \sum_{m=k-1}^{\infty}\left(\frac{\mathrm{P}(L_m = k-1)\log m}{(\log\log m)^{k-3}}\right)^2 \frac{(\log\log m)^{2(k-3)}}{m\log^2 m}$$
$$\leqslant C\sum_{m=k-1}^{\infty} \frac{(\log\log m)^{2(k-3)}}{m\log^2 m} < \infty.$$

And note that

$$\lim_{n\to\infty} \mathrm{P}\big(\eta_n > (1+\varepsilon)\log n\big) = 0,$$

then (4.28) follows. Therefore,

$$\frac{(k-2)!\log n}{(\log\log n)^{k-2}} \cdot \sum_{m=k-1}^{n-1} \mathrm{P}(L_m = k-1)\mathrm{E}\frac{\zeta_{n,m}}{\eta_n}$$
$$\geqslant \frac{(k-2)!\log n}{(\log\log n)^{k-2}} \cdot \sum_{m=k-1}^{n-1} \mathrm{P}(L_m = k-1)\mathrm{E}\left(\frac{\zeta_{n,m}}{\eta_n} I\big(\eta_n \leqslant (1+\varepsilon)\log n\big)\right)$$
$$\geqslant \frac{1}{1+\varepsilon}\frac{(k-2)!}{(\log\log n)^{k-2}} \cdot \sum_{m=k-1}^{n-1} \mathrm{P}(L_m = k-1)\mathrm{E}\big(\zeta_{n,m} I\big(\eta_n \leqslant (1+\varepsilon)\log n\big)\big)$$
$$= \frac{1}{1+\varepsilon}\frac{(k-2)!}{(\log\log n)^{k-2}} \cdot \sum_{m=k-1}^{n-1} \mathrm{P}(L_m = k-1)\mathrm{E}\zeta_{n,m}$$
$$- \frac{1}{1+\varepsilon}\frac{(k-2)!}{(\log\log n)^{k-2}} \cdot \sum_{m=k-1}^{n-1} \mathrm{P}(L_m = k-1)\mathrm{E}\big(\zeta_{n,m} I\big(\eta_n > (1+\varepsilon)\log n\big)\big).$$

Thus, by (4.25) and (4.27),

$$\liminf_{n\to\infty} \frac{(k-2)!\log n}{(\log\log n)^{k-2}} \cdot \sum_{m=k-1}^{n-1} \mathrm{P}(L_m = k-1)\mathrm{E}\frac{\zeta_{n,m}}{\eta_n} \geqslant \frac{1}{1+\varepsilon}.$$

Since $\varepsilon > 0$ is arbitrary, (4.26) holds. The proof of the theorem is completed. □

Remark 4.1. In other words, for any nature number $k \geqslant 2$, as $n \to \infty$,

$$\mathrm{P}(L_n = k) \sim \frac{(\log\log n)^{k-2}}{(k-2)!\log n}.$$

Note that

$$\sum_{k=2}^{\infty} \frac{(\log\log n)^{k-2}}{(k-2)!\log n} = \frac{1}{\log n} \sum_{k=0}^{\infty} \frac{(\log\log n)^k}{k!} = 1,$$

which shows that the result is reasonable.

References

[1] E. Bach, Random bisection and evolutionary walks, J. Appl. Probab. 38 (2001) 582–596.

[2] F. Bergeron, P. Flajolet, B. Salvy, Varieties of increasing trees, in: J.C. Raoult (Ed.), Proc. 17th Coll. Trees in Algebra and Programming, in: Lecture Notes Comput. Sci., vol. 581, Springer, Berlin, 1992, pp. 24–48.

[3] L. Devroye, Applications of the theory of records in the study of random trees, Acta Inform. 26 (1988) 123–130.

[4] Q.Q. Feng, C. Su, Z.S. Hu, Branching structure of uniform recursive trees, Sci. China, submitted for publication.

[5] J. Gastwirth, A probability model of a pyramid scheme, Amer. Statist. 31 (1977) 79–82.

[6] C.A. Macken, P.S. Hagan, A.S. Perelson, Evolutionary walks on rugged landscapes, SIAM J. Appl. Math. 51 (1991) 799–827.

[7] C.A. Macken, A.S. Perelson, Protein evolution on rugged landscapes, Proc. Nat. Acad. Sci. USA 86 (1989) 6191–6195.

[8] C.A. Macken, A.S. Perelson, Affinity maturation on rugged landscapes, in: A. Perelson, S. Kauffman (Eds.), Molecular Evolution on Rugged Landscapes, Addison–Wesley, Reading, MA, 1991.

[9] H.M. Mahmoud, Limiting distributions for path lengths in recursive trees, Probab. Engrg. Inform. Sci. 5 (1991) 53–59.

[10] H.M. Mahmoud, Distances in random plane-orients recursive trees, J. Comput. Appl. Math. 41 (1992) 237–245.

[11] A. Meir, J.W. Moon, Climbing certain types of rooted trees. II, Acta Math. Acad. Sci. Hungar. 31 (1978) 43–54.

[12] A. Meir, J.W. Moon, On the altitude of nodes in random trees, Canad. J. Math. 30 (1978) 997–1015.

[13] J.W. Moon, The distance between nodes in recursive trees, in: London Math. Soc. Lecture Note Ser., vol. 13, Cambridge Univ. Press, London, 1974, pp. 125–132.

[14] D. Najock, C. Heyde, On the number of the terminal vertices in certain random trees with an applications to stemma construction in philology, J. Appl. Probab. 19 (1982) 675–680.

[15] J. Szymański, On the maximum degree and the height of a random recursive tree, in: M. Karoński, A. Ruciński (Eds.), Random Graphs'87, Wiley, North-Holland, 1990, pp. 313–324.

[16] J.M. Xu, Graph Theory and its Applications, Press of University of Science and Technology of China, 1998.